

Language detection via Accept-Language

Mikhail Popov (Analysis and Reporting), Erik Berhardson (Engineering and Data Collection), Dan Garry (Management), Oliver Keyes (Review)

January 11, 2016

Summary

This was an A/B test to determine the impact of switching search languages in the case that a query produces zero results or fewer than 3 results. While we previously used the elasticsearch plug-in “langdetect” to detect the query’s language, we suspected that we should first try finding a valid language using the Accept-Language header, if it is detected.

We found evidence that the Accept-Language header detection makes a slight positive difference to the zero results rate, with 3.18-3.34% more requests getting some results in the test group than the control group, and the test group being 1.026-1.029 more likely to get some results than the control group when we found a valid language via Accept-Language detection.

Introduction

Currently, if a user searches the German Wikipedia for a query written in Russian, they probably will not get any results back. (There are instances where they would.) So we decided that in the case where the user does not get any results back, perhaps we should try detecting the language of their query and matching it up with the correct language wiki.

We previously tested the “langdetect” plugin for elasticsearch and [saw a significant decrease](#) in the zero results rate.

We suspected that using the first non-English Accept-Language HTTP header will provide a good proxy for the language the query is in when the query returns no results against the wiki it was already run against. We suspected that this is a better proxy than the existing elasticsearch “langdetect” plugin.

There were three total groups of requests: the control group with status quo settings (wherein no language switching is done), the requests for which we performed a language detection based on Accept-Language (AL) header and then elasticsearch plug-in (es-plugin) when there were 0 results returned, and the requests for which we performed a language detection when there were less than 3 results returned. If the AL detection could not find a valid language, we fell back on the es-plugin detection.

Methods

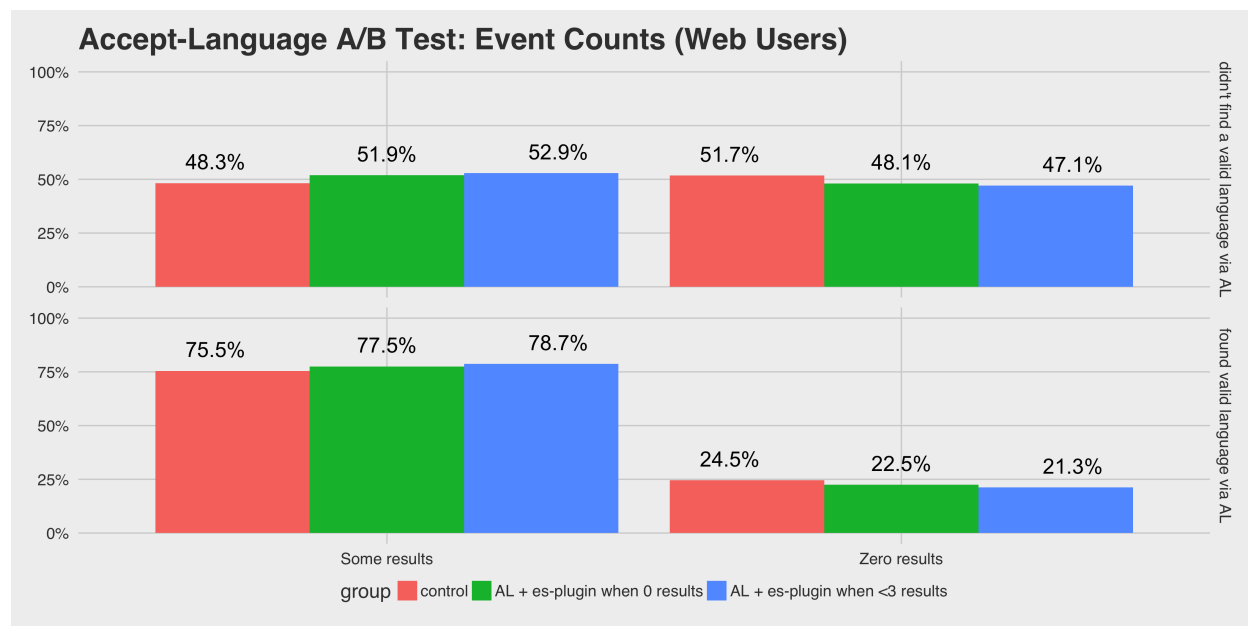
The sampling process was random, but on a per-user basis; if a user was selected, all of their queries were included. One out of 7 requests were sampled; thus, each group had a sampling rate of 1 in 21 sample requests.

First, we counted the number of events per user (usually uniquely identified by their IP, User Agent, and X-Forwarded-For, but with exceptions). After looking at the counts of events, we restricted our **Web** dataset to users who made less than 7 search requests, which accounts for 99% of the data. The top 1% of **Web** users was excluded from analysis because they would skew the results.

While API requests are included in this test, the API user must pass an explicit flag opting into the general query rewriting feature which basically no-one enables. Any measurable effect will be entirely within users of the **Web**-based search.

To perform the analysis, we employed Agresti and Min’s (2005) methods for computing Bayesian confidence intervals.

Results



We compared controls with the group where AL & es-plugin were used when 0 results were returned, and computed 95%-probability intervals. Among those for who we *were not able to detect a valid language via AL*, 2.8%-4.4% more users received results in the test group and were 1.058-1.093 times more likely to receive results than the control group. Among those for who we *were able to detect a valid language via AL*, 2%-2.2% more users received results in the test group and were 1.026-1.029 times more likely to receive results than the control group.

We also performed this test with a test group who had received **less than 3 results**. Among those for who we *were not able to detect a valid language via AL*, 3.87%-5.46% more users received results in the test group and were 1.08-1.11 times more likely to receive results than the control group. Among those for who we *were able to detect a valid language via AL*, 3.18-3.34% more users received results in the test group and were 0.95-1.044 times more likely to receive results than the control group, which is to say we cannot conclude that their search results were impacted in a significantly positive way.

Conclusion/Discussion

The Accept-Language header detection makes a slight positive difference to the zero results rate, with 3.18-3.34% more requests getting some results in the test group than the control group, and the test group being 1.026-1.029 more likely to get some results than the control group when we found a valid language via Accept-Language detection.

As with the previous test, we recommend collecting data about users' interaction with the results to see whether the language detection (via AL and/or es-plugin) produces useful results, not just some results. What is the point of giving users results if the results we're giving them might be nonsensical?

References

- [User:EBernhardson \(WMF\)/Notes/Accept-Language](#)
- [Discovery/Testing: Language detection via Accept-Language](#)
- Agresti, A. and Min, Y. (2005). Frequentist performance of Bayesian confidence intervals for comparing proportions in 2x2 contingency tables, *Biometrics*, **61**, 515-523. [doi:10.1111/j.1541-0420.2005.031228.x](https://doi.org/10.1111/j.1541-0420.2005.031228.x)